

## ⑫ 公開特許公報(A)

昭62-253200

⑤ Int. Cl.

G 10 L 9/00  
9/02  
9/14

識別記号

庁内整理番号

M-8622-5D  
L-8622-5D  
M-8622-5D

④ 公開 昭和62年(1987)11月4日

審査請求 有 発明の数 1 (全5頁)

⑥ 発明の名称 音声合成器

⑦ 特 願 昭61-295829

⑧ 出 願 昭61(1986)12月11日

優先権主張 ⑨ 昭61(1986)1月28日 ⑩ 日本(JP) ⑪ 特願 昭61-17153

⑫ 発 明 者 田 口 哲 東京都港区芝5丁目33番1号 日本電気株式会社内  
 ⑬ 出 願 人 日本電気株式会社 東京都港区芝5丁目33番1号  
 ⑭ 代 理 人 弁理士 内 原 晋

## 明 細 書

1. 発明の名称  
音声合成器

## 2. 特許請求の範囲

ホルマントにもとづいて音声を作成する音声合成器において、

音声信号の変化のゆるやかな区間もしくは連続的区間はパラメータ補間をホルマント領域で実施し、音声信号の変化の激しい区間もしくは非連続的区間はパラメータ補間を隠形予測係数領域で実施するパラメータ補間手段を備えて成ることを特徴とする音声合成器。

## 3. 発明の詳細な説明

〔産業上の利用分野〕

本発明は音声合成器に関し、特にパラメータ補間処理効率の改善を図った音声合成器に関する。

〔従来の技術〕

ホルマントはパラメータ補間特性が良好とされており、このため一般的にホルマント領域でのパラメータ補間が行なわれることが多い。

しかしながら、語頭、渡り等の音声信号の部分ではホルマント(または極)は隣接する分析フレーム間で非連続的となることが多く、このような区間を補間すると聴覚的に極めて不自然な合成音が発生することが経験的にもよく知られている。

このような問題に対して、従来、パラメータの非連続的区間ではパラメータの分析窓長を可変とする等の手段を取り入れ、変化の激しい区間の分析窓時間長を短かくし、この区間では補間を禁止することによって不自然な合成音の発生の排除を図っている。

〔発明が解決しようとする問題点〕

上述した従来の音声合成器は、合成におけるパラメータ補間制御等にもとづいて本質的に処理時間が増大し合成に大きな遅延が発生するという欠点がある。

本発明の目的は上述した欠点を除去し、音声信

号の変化の緩急に対応してパラメータ補間をホルマント領域と線形予測係数領域とに分けて実施するという手段を備えることにより、合成音の品質を著しく改善するとともに合成時の遅延も大幅に改善した音声合成器を提供することにある。

〔問題点を解決するための手段〕

本発明の装置は、ホルマントにもとづいて音声合成する音声合成器において、音声信号の変化のゆるやかな区間もしくは連続的な区間はパラメータ補間をホルマント領域で実施し、音声信号の変化の激しい区間もしくは非連続的な区間はパラメータ補間を線形予測係数で実施するパラメータ補間手段を備えて構成される。

〔実施例〕

次に図面を参照して本発明を詳細に説明する。

第1図は本発明の一実施例のブロック図である。第1図に示す一実施例は、ホルマント/αパラメータ変換器1、α/K変換器2、補間制御器3、補間器4、K合成フィルタ5、D/Aコンバータ6、可変利得増幅器7、切替器8、ピッチパルス

- 3 -

と替ける。従ってホルマント周波数は

$$f_i = \frac{\omega_i}{2\pi} = \frac{1}{2\pi} \cdot \frac{1}{T} \arg(Z_i) \text{ (Hz)} \quad \dots\dots\dots(2)$$

(ただしTは標準化周期)

ホルマント帯域幅は

$$b_i = \frac{1}{\pi} \cdot \frac{1}{T} |\log r_i| \text{ (Hz)} \quad \dots\dots\dots(3)$$

と求められる。

尚、上記αパラメータとホルマントとの関係は「音声情報処理の基礎」、斉藤収三、中田和男著、オーム社発行の第6章3項に詳しく述べられている。

さて、入力したホルマントパラメータの個数をP/2個対応するホルマントの周波数と帯域幅を{f<sub>i</sub>, b<sub>i</sub>}, i=1, 2, ...,  $\frac{P}{2}$  とすると、(2)式より

$$\omega_i = 2\pi f_i \quad \dots\dots\dots(3)$$

(3)式より

$$r_i = e^{b_i T} \quad \dots\dots\dots(4)$$

- 5 -

発生器9、雑音発生器10等を備えて構成される。

一定分析周期ごとに抽出されるホルマントパラメータが、ホルマント/αパラメータ変換器1と補間制御器3に供給される。この場合、ホルマントパラメータは分析フレームごとに抽出された各ホルマント周波数、ならびにその帯域幅に関するデータである。

ホルマント/αパラメータ変換器1は、入力したホルマントパラメータを以下の方法で線形予測係数、即ちαパラメータに変換する。

αパラメータを{α<sub>i</sub>}, i=1, 2, ..., Pとするとホルマントは下記(1)式の根{Z<sub>i</sub>}, i=1, 2, ..., Pより導出される。

$$Z^P + \alpha_1 Z^{P-1} + \alpha_2 Z^{P-2} + \dots + \alpha_{P-1} Z + \alpha_P = 0 \quad \dots\dots(1)$$

(1)式は実係数のP次方程式であり、一般にP/2個の共役複素根をもつ。その共役根をZ<sub>i</sub>=r<sub>i</sub>e<sup>jθ<sub>i</sub></sup>,  $\bar{Z}_i = r_i e^{-j\theta_i}$ と表わす。

一方

$$Z = e^{sT} = e^{(-a + j\omega)T} = e^{-aT} e^{j\omega T} = r e^{j\theta}$$

- 4 -

故に

$$a_i = -b_i \pi T \quad \dots\dots\dots(5)$$

αパラメータは下記(6)式により求められる。

$$\begin{aligned} & Z^P + \alpha_1 Z^{P-1} + \alpha_2 Z^{P-2} + \dots + \alpha_{P-1} Z + \alpha_P \\ &= \{(Z - e^{(-a_1 + j\omega_1)T})(Z - e^{(-a_1 - j\omega_1)T})\} \dots\dots \\ &\dots\dots \{(Z - e^{(-a_{\frac{P}{2}} + j\omega_{\frac{P}{2}})T})(Z - e^{(-a_{\frac{P}{2}} - j\omega_{\frac{P}{2}})T})\} \dots\dots(6) \end{aligned}$$

算出されたαパラメータ{α<sub>i</sub>}はα/K変換器2に出力される。

α/K変換器2は入力したαパラメータ{α<sub>i</sub>}を「Linear Prediction of Speech」, J. D. Markel, A. H. Gray, Jr, Springer-Verlag, 5, 2, 2

「Step-down Procedure」に示される(5, 12)式の関係を利用してKパラメータ{K<sub>i</sub>}, i=1, 2, ..., Pを求める。尚、(5, 12)式においてα<sub>m</sub>, iと示された係数がα<sub>i</sub>に対応し、m=Pである。又、K<sub>m</sub>はK<sub>P</sub>に対応する。

α/K変換器2は求めたKパラメータ{K<sub>i</sub>}を補間制御器3へ供給する。

- 6 -

補間制御器3は、入力するホルマントパラメータを予め設定する個数の分析フレーム単位で監視しつつ、ホルマントパラメータ相互間の変動が予め設定する一定の判定基準内に入っているか否かの判定を行ない、変動が判定基準内におさまっており従って音声信号の変化がゆるやかな分析フレーム区間、もしくは連続的と考えられる分析フレーム区間は音声合声フィルタの係数とする特徴パラメータに対する補間処理はホルマント領域で実施するものとし、ホルマントパラメータを出力ライン401を介して補間器4に供給する。なお、上述した判定基準は、多くの音声資料にもとづいて予め設定される。本実施例では下記(7)式を利用して判定が行なわれる。

$$J = 0, \text{ if } \sum_{i=1}^{\frac{P}{2}} \{ |f_i^{(B)} - f_i| + U_i | b_i^{(B)} - b_i | \} < \nu$$

$$J = 1, \text{ if } \sum_{i=1}^{\frac{P}{2}} \{ |f_i^{(B)} - f_i| + U_i | b_i^{(B)} - b_i | \} \geq \nu$$

$$i = 1, 2, \dots, \dots (7)$$

ここで、 $J = 0$  の場合、変動が基準内である事

- 7 -

る。

$$\left. \begin{aligned} f_i^{(1)} &= \{ T_x f_i^{(R)} + (T_f - T_x) f_i \} / T_f \\ b_i^{(1)} &= \{ T_x b_i^{(R)} + (T_f - T_x) b_i \} / T_f \end{aligned} \right\} \dots (8)$$

$$i = 1, 2, \dots, \frac{P}{2}$$

ただし  $f_i^{(R)}$ ,  $b_i^{(R)}$  は現フレームに隣接する過去のフレームのホルマントデータである。

補間制御器3は、こうして入力したホルマントパラメータ補間データを出力ライン301を介してホルマント/ $\alpha$ パラメータ変換器1に供給し、ホルマント/ $\alpha$ パラメータ変換器1はこれを $\alpha$ パラメータに変換したあと $\alpha/K$ 変換器2に出力する。 $\alpha/K$ 変換器2では $K$ パラメータに変換されて補間制御器3に供給されるが、こうしてホルマントパラメータの補間にもとづいて得られた $K$ パラメータは出力ライン302を介して $K$ 合成フィルタ5にそのフィルタ係数として提供される。

こうして、音声信号のゆるやかな区間もしくは連続的な区間はパラメータ補間をホルマント領域での補間特性の良さが発揮されるよう補間制御器

- 9 -

を、 $J = 1$  の場合、変動が基準外である事を示す。又、 $f_i^{(B)}$ ,  $b_i^{(B)}$ ,  $i = 1, \frac{P}{2}$  は過去の分析フレームのホルマントデータである。無論過去の分析フレームとして、現分析フレームと相隣接するフレームに限定する事も、過去数フレームを対象とする事も可能である。

又、ホルマントデータの時間的変動の激しい部分では短時間平均電力の時間的変動も激しいという経験的事実を利用して、このような判定基準に代えて分析フレームごとに得られる短時間平均電力の如き電力パラメータを利用して判定することも可能である。

補間器4は、入力したホルマントパラメータに対し所定の形式の補間処理、本実施例の場合は直線補間を実施しこの補間データを入力ライン402を介して補間制御器3に供給する。

無論、この補間は各ホルマント周波数、帯域幅毎に独立に実施されるものである。例えばフレーム周期を $T_f$ とすると現フレームより $T_x$ だけ過去のデータ  $f_i^{(1)}$ ,  $b_i^{(1)}$  は下記(8)式を利用して算出され

- 8 -

3による補間制御が行なわれる。

さて、補間制御器3に入力するホルマントパラメータ相互間の変動が一定の判定基準から逸脱したときは補間制御器3による次の補間制御が行なわれる。このような状態は、音声の語頭、渡りの部分等であり、ホルマント（または極）が隣接分析フレーム間で非連続になり易くその結果もたらされるものである。このような区間もホルマント領域補間を施すと聴覚的に極めて不自然な合成音が発成し易いことが多くの事例での経験的知識としてよく知られている。このような区間ではむしろ線形予測係数領域補間の方が合声音質の不自然さははるかに少なく、たとえば線形予測係数として $K$ パラメータを利用する $K$ パラメータ領域補間を変化の激しい区間もしくは非連続的区間に適用することによって良好な合成音とすることが可能となる。

補間制御器3は、入力するホルマントパラメータを監視しつつ、ホルマントパラメータ相互間の変動が予め設定する一定の判定基準からはみ出し

- 10 -

たと判定した区間では $\alpha/K$ 変換器2の出力するKパラメータを補間対象として補間を実施するものとし、出力ライン401を介してKパラメータを補間器4に供給、これに所定の形式の補間処理を施さしめ、補間データとして入力ライン402を介してこれを取込みフィルタ係数として出力ライン302を介してK合成フィルタ5に提供する。

このKパラメータでの補間はホルマント補間と同様に下記(9)式を利用して行なわれる。

$$K_i^{(1)} = \{T_x K_i^{(R)} + (T_f - T_x) K_i\} / T_f \quad \dots\dots (9)$$

$$(i = 1, 2, \dots, P)$$

ただし $K_i^{(1)}$ は補間により求められたKパラメータ、 $K_i^{(R)}$ は現フレームに隣接する過去のフレームのKパラメータである。

補間制御器3は、こうして、入力するホルマントパラメータを一定の判定基準と比較、照合しつつ、入力が判定基準内に収まるか否かでホルマント領域か線形予測係数領域かの選択を行ないいずれかの領域での補間を実施するが、この場合、一定

-11-

もしくは雑音発生器10のいずれかを入力、増幅し、これを音源信号としてK合成フィルタ5に提供する。

ピッチパルス発生器9は、分析側等から提供されるピッチ周期データに対応する繰返し数のピッチパルス列を分析フレームごとに発生しこれを切替器8に供給する。

また、雑音発生器10は白色雑音を発生しこれを切替器8に供給する。

切替器8は分析フレームごとの有声/無声情報を分析側等から受け、分析フレームごとの音声信号が有声を指定するときにはピッチパルス発生器9の出力を、また無声を指定するときには雑音発生器10の出力を可変利得増幅器7に出力するように切替え、こうして可変利得増幅器7からは擬似の音源信号が分析フレームごとにK合成フィルタ5に提供されてこれを駆動し、K合成フィルタ5からはデジタル形式の音声信号が合成されてD/Aコンバータ6に供給される。

D/Aコンバータ6は、こうして入力したディ

-13-

の判定基準と入力ホルマントパラメータ列との比較照合は、前記(7)式を利用する外、たとえば、入力ホルマントパラメータ値にもとづき得られる最小自乗誤差の比較、もしくはあらかじめ設定するモデル関数との非線形時間軸変換を介しての比較、さらには前述した電力スペクトルの比較等種類の手法にもとづく比較照合が利用できる。

こうして音声信号の変化の緩急性をらびに連続性の程度を判定し、これが判定基準から逸脱するときには補間制御器3は補間器4に供給するパラメータをホルマントパラメータからKパラメータに切替えるように補間制御を行なう。

K合成フィルタ5はフィルタ係数にKパラメータを利用する音声合成フィルタであり補間制御器3を介して補間器4からKパラメータをフィルタ係数として受けつつ可変利得増幅器7の出力する音源信号によって駆動される。

可変利得増幅器7は電力パラメータとしての短時間平均電力に対応してその利得を可変としつつ、有声、無声情報に対応してピッチパルス発生器9

-12-

デジタル再生音声信号をアナログ量に変換したりえ所定の遮断周波数の低域通過フィルタを通し所望の音声信号として出力する。

こうして良好な合成音がほぼリアルタイムで、合成のための遅延時間を大幅に圧縮した状態で再生できる。

なお、本実施例では、短時間平均電力、有声/無声情報、ピッチ周期データにもとづいて発生した擬似音源信号を利用して音声合成フィルタを駆動しているが、この代りに他の方法で発生した音源情報、たとえばマルチパルス励振や残差励振型の形式による音源情報を利用するものとしてもよい。

(発明の効果)

以上説明したように本発明によれば、音声信号の変化の緩急に対応して、変化のゆるやかな区間もしくは連続的区間はパラメータ補間をホルマント領域で実施し、変化の激しい区間もしくは非連続的区間はパラメータ補間を線形予測係数領域で実施するという手段を備えて音声の合成を図ること

-14-

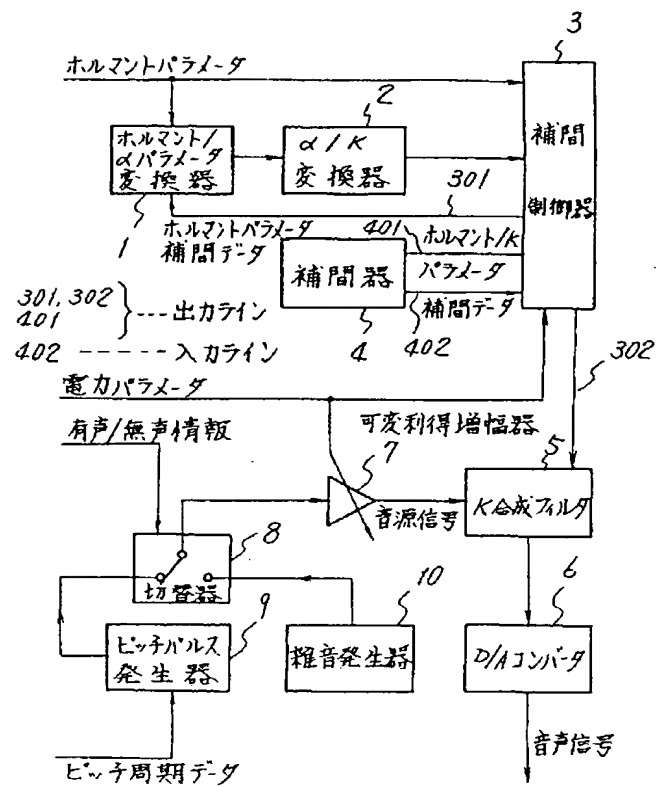
とにより、合成音声の品質を著しく改善するとともに合成時の遅延も大幅に減少しうる音声合成器が実現できるという効果がある。

#### 4. 図面の簡単な説明

第1図は本発明の一実施例を示すブロック図である。

1……ホルマント/ $\alpha$ パラメータ変換器、2…… $\alpha$ / $K$ 変換器、3……補間制御器、4……補間器、5…… $K$ 合成フィルタ、6……D/Aコンバータ、7……可変利得増幅器、8……切替器、9……ピッチパルス発生器、10……雑音発生器。

代理人 井理士 内 原 晋



第1図